Molecular data show conserved DNA locations distinguishing lung cancer subtypes and regulation of immune genes

Hurkmans, Daan P; Tamminga, Menno; van Es, Bram; Peters, Tom; Karman, Wouter; van Wijck, Rogier T A; van der Spek, Peter J; Tauber, Tjebbe; Los, Maureen; van Schetsen, Anouk

Published in:
Lung Cancer

DOI:
10.1016/j.lungcan.2020.06.008

Publication date:
2020

Lung Cancer 146 (2020) 341–349

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Molecular data show conserved DNA locations distinguishing lung cancer subtypes and regulation of immune genes

Daan P. Hurkmans[a,1], Menno Tamminga[b,1], Bram van Es[c], Tom Peters[d], Wouter Karman[d], Rogier T.A. van Wijck[a], Peter J. van der Spek[a], Tjebbe Tauber[e], Maureen Los[d], Anouk van Schetsen[d], Thu Vu[d], T. Jeroen. N. Hiltermann[b], Ed Schuuring[b], Joachim G.J.V. Aerts[a], Sissy Chen[d], Harry J.M. Groen[b,*]

[a] Erasmus University Medical Center, Departments of Pulmonary Diseases, Internal Medicine and Pathology, Bioinformatic Unit, Dr. Molewaterplein 40, 3015 GD, the Netherlands
[b] University of Groningen and University Medical Center Groningen, Departments of Pulmonary Diseases and Pathology and Medical Biology, Hanzeplein 1, 9713 GZ, Groningen, the Netherlands
[c] Otravo B.V., Suikersilo-West 41, 1165 MP, Amsterdam-Halfweg, the Netherlands
[d] PricewaterhouseCoopers Advisory NV, Thomas R. Malthusstraat 5, 1066 JR, Amsterdam, the Netherlands
[e] ABN-AMRO, Foppingadreef 22, 1102 BS Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Methylation
DNA
Immune gene
Microenvironment
Lung cancer
NSCLC
Adenocarcinoma
Squamous cell carcinoma

ABSTRACT

Introduction: Non-small-cell lung cancer exhibits a range of transcriptional and epigenetic patterns that not only define distinct phenotypes, but may also govern immune related genes, which have a major impact on survival.
Methods: We used open-source RNA expression and DNA methylation data of the Cancer Genome Atlas with matched non-cancerous tissue to evaluate whether these pretreatment molecular patterns also influenced genes related to the immune system and overall survival.
Results: The distinction between lung adenocarcinoma and squamous cell carcinoma are determined by 1083 conserved methylation loci and RNA expression of 203 genes which differ for > 80 % of patients between the two subtypes. Using the RNA expression profiles of 6 genes, more than 95 % of patients could be correctly classified as having either adeno or squamous cell lung cancer. Comparing tumor tissue with matched normal tissue, no differences in RNA expression were found for costimulatory and co-inhibitory genes, nor genes involved in cytokine release. However, genes involved in antigen presentation had a lower expression and a wider distribution in tumor tissue.
Discussion: Only a small number of genes, influenced by DNA methylation, determine the lung cancer subtype. The antigen presentation of cancer cells is dysfunctional, while other T cell immune functions appear to remain intact.

1. Introduction

Smoking induced lung cancer has a large number of DNA mutations while other environmental factors such as air pollution may cause a different distribution in DNA mutations, often observed in non-smokers in genes like *EGFR*, *BRAF*, *HER2* and *ALK* [1]. Lung cancer is traditionally subdivided into small-cell lung cancer and non-small-cell lung cancer (NSCLC) with the latter being divided into two main subtypes, adenocarcinoma and squamous cell carcinoma (SCC).

It is known that DNA methylation is affected by age, smoking, emphysema and histological subtype [2]. Changes in the methylation pattern affects RNA expression, leading not only to different

* Corresponding author at: Department of Pulmonary Diseases, University Medical Center Groningen, Hanzeplein 1, P.O. Box 30.0001, 9700 RB Groningen, the Netherlands.
E-mail addresses: d.hurkmans@erasmusmc.nl (D.P. Hurkmans), m.tamminga@umcg.nl (M. Tamminga), bramiozo@gmail.com (B. van Es), tom.peters@pwc.com (T. Peters), wouter.karman@pwc.com (W. Karman), r.vanwijck@erasmusmc.nl (R.T.A. van Wijck), p.vanderspek@erasmusmc.nl (P.J. van der Spek), tjebbe.tauber@nl.abnamro.com (T. Tauber), maureen.los@pwc.com (M. Los), anouk.van.schetsen@pwc.com (A. van Schetsen), thu.vu@pwc.com (T. Vu), t.j.n.hiltermann@umcg.nl (T.J.N. Hiltermann), e.schuuring@umcg.nl (E. Schuuring), j.aerts@erasmusmc.nl (J.G.J.V. Aerts), sissy.chen@pwc.com (S. Chen), h.j.m.groen@umcg.nl (H.J.M. Groen).
[1] Authors contributed equally.

https://doi.org/10.1016/j.lungcan.2020.06.008
Received 20 March 2020; Received in revised form 3 June 2020; Accepted 9 June 2020

phenotypes, but also to a different effect on the immune activation. This may not only be reflected in the expression of human leukocyte antigen (HLA) or PD-L1 on tumor cells, but also in the tumor microenvironment. The Cancer Genome Atlas group (TCGA) performed molecular studies on lung adenocarcinoma and SCC identifying driver oncogenes and loss-of-function mutations in the HLA-A class I major histocompatibility gene [3,4]. Molecular classification approaches were made by clustering phenotypes on different platforms [5]. In a more recent study, a "cluster-of-clusters" analytic approach on differential DNA expression showed three distinct subtypes within SCC and six within adenocarcinomas [6]. Three of the adenocarcinoma subtypes had high expression of several immune related genes including PD-L1, PD-L2, CD3 and CD8.

To date, the role of epigenetic modifications in relation to tumor responses in NSCLC remains to be clarified. Global DNA hypomethylation at repeated sequences has been identified in tumor cells in combination with DNA hypermethylation at specific loci [7]. CpG dinucleotides are highly represented in repeated sequences of the genome (LINE, SINE) and in the promotor regions of about 65 % of the human genome and it is suggested that these play a role in regulation of gene expression. Biopsies of both adenocarcinoma and SCC show differences in methylation and in immune infiltrate [8]. Although their tumor response to checkpoint inhibitors is similar (1 year progression-free survival of 21 % and 19 % for respectively SCC and non-SCC in the Checkmate studies), different determinants driving tumor response and resistance may be involved [9]. We hypothesized firstly, that a clear separation based on DNA methylation and RNA expression can be made between the NSCLC subtypes, and secondly, that methylation controls immune modulating genes in the tumor DNA and tumor-intrinsic defects must be present.

2. Results

2.1. Methylation in NSCLC

We used 1024 unique patients from NSCLC TCGA datasets. Based on DNA methylation a PCA analysis appeared to be capable of separating adenocarcinoma and SCC (Fig. 1a). The prediction model based on the PCA outcomes correctly identified all but one of the included patients as either adenocarcinoma or SCC. Importantly, stratification for high and low purity (indicated as the proportion of tumor cell content) of the processed samples did not influence the findings. Remarkably, of all methylation probes with a ks-score ≥ 0.95, the mean corrected differences ranged from $+0.02$ to $+0.15$ (scale -1 - +1), implicating a very small variation in methylation for these highly conserved loci in both phenotypes and a consistent stronger methylation of adenocarcinoma compared to SCC (Fig. 2).

After we observed significant differences in methylation probe distribution between the subtypes, we continued with a best split analysis. The algorithm identified 1083 methylation probes (out of a total of 485,577 probes) which individually could be used to correctly classify at least 85 % of patients. Next, we looked into the chromosomal position of the different methylated probes. An even distribution along the genome was determined at increasingly stringent ks-scores (Fig. 3a,b and Extended Data Table 1).

The higher ks-score indicates a better separability between histologic subtypes (Supplementary Information). The methylation pattern for each chromosome characterized by individual probes with ks-score $\geq$ 95 % was distributed over all chromosomes except the X-chromosome. Probes with a high ks-score over 97 % for separability represented conserved methylated loci that preserve the difference between phenotypes.

To address morphological differences between adenocarcinoma and SCC based on DNA methylation, an enrichment analysis was performed on the gene level based on the methylation probes that are most distinct for phenotype (ks-score > 0.95; n = 2101 mapped genes). Remarkably, these and other genes showed a low methylation rate compared to normal tissue. The main canonical pathways that are most distinct for NSCLC subtypes are DNA repair pathways (Extended Data Fig. 1). However, based on the relative methylation of these genes, mechanisms involved in response to the category "viral infections" (z-score 10.7, $p < 0.001$;) were more activated in SCC compared to adenocarcinoma, whereas mechanisms involved in cell death (z-score -17.8, $p < 0.001$) were inhibited. Central genes involved in "viral infection" that are found to be differential in SCC compared to adenocarcinoma include *IRF3*, *NFKB1*, *RELA* (also known as *NFKB3*), *STAT3*, *SRPK1* and *TRIM*.

2.2. Expression in NSCLC

We next aimed to investigate to what extent methylation influences RNA expression levels. We observed that DNA methylation explains approximately 40–55 % of the inversely correlated variation in the RNA expression. This percentage, however, is not only dependent on the β-value level that would biologically lead to effective epigenetic gene suppression, but also on the correlation between gene expression and methylation (Fig. 4e). Approximately 60 % of methylation probes (different to the random 50 %) are inversely correlated with RNA expression by selecting only those methylation probes that are assumed to have an epigenetic suppressive effect (average β-value > 0.25) on DNA transcription and with a significant correlation (correlation coefficient > 0.5 or < -0.5). Of note, methylation probes with a negative, positive, or no relationship with gene expression could be determined as illustrated in Fig. 4b–d (Supplementary information, sub 5). In general, genes that were heavily methylated showed a lower RNA expression than genes that had a moderate or low methylation rate.

Unsupervised principal component analysis of the transcriptome led to a slightly less accurate separation of NSCLC phenotypes (Fig. 1b) than that based on DNA methylation data (Fig. 1a). Comparing tumor with non-cancerous tissue confirms that the RNA expression pattern is specific for NSCLC (Fig. 1b). The best split analysis on RNA data identified 203 genes, of which the expression was different in 85 % of cases between SCC and adenocarcinoma. Of these, differences in RNA expression of five genes (*KRT5, DSC3, DSG3, TP63, CALML3*), and one miRNA (*MIR205HG*) combined could separate both subtypes with an accuracy of 95 %.

2.3. Bilevel molecular analysis in NSCLC

We selected the top 500 methylation probes with their corresponding genes and the top 500 genes based on RNA expression and found an overlap of 41 genes related to the separation of the NSCLC phenotypes based on both DNA methylation and RNA expression (Extended Data Table 2). Gene enrichment analysis by Ingenuity Pathway Analysis revealed that *TP63* was an important upstream regulator, with elevated expression in SCC compared to adenocarcinoma. Target molecules of *TP63* in the list of 41 genes included *CSTA, SNAI2, DST, ACTL6A, KRT7* and the miRNA *MIR205HG*, and their expression was in the same predicted direction as the *TP63* activation in SCC.

2.4. Immune modulating genes and methylation

As expected, methylation was inversely correlated for most methylation probes, e.g. higher level of *HLA-B*, *TAP1*, *CD2* methylation leads to lower RNA expression (Fig. 4a). Of note, none of the 1083 methylation probes identified by the best split analysis for histological subtype included any of the selected immune related genes.

We performed a principal component analysis including all RNA expressing genes and determined the weight of each component of immune modulatory gene groups: T cell co-inhibitory (COINHIB), T cell co-stimulator (COSTIM), T cell antigen presentation (AGPRES) and T cell cytokines/chemokines (CYTOCHEM) (Fig. 5 and Extended Data Fig. 2). Genes in the immune related groups were gathered in two

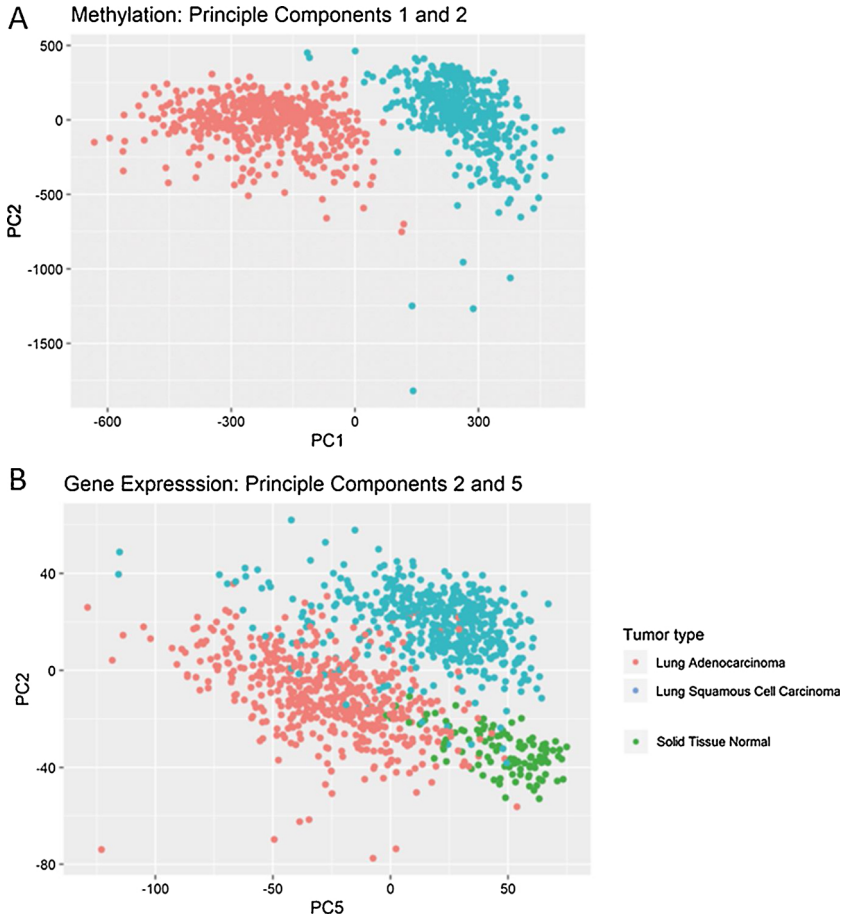

Fig. 1. Clustering of main lung cancer phenotypes based on DNA methylation and RNA expression.
A) Pulmonary adenocarcinoma is distinguished from squamous cell carcinoma by DNA methylation and B) to a slightly lesser extent by RNA expression.

clusters, in PC3 and PC6 − 9. Antigen presenting gene expression was positively associated with co-stimulatory gene expression, not only in tumor but also in non-cancerous tissue (Extended Data Fig. 2b, c).

Genes involved in the immune response towards endogenous retroviral sequences were also more methylated in adenocarcinoma compared to SCC, but we did not study the repetitive DNA areas with high versus low methylation.

2.5. Antigen presenting gene expression in tumor

Average RNA expression of genes in the antigen presenting gene group was lower in tumor than in non-cancerous tissue and showed a larger variation (SD) in expression (Fig. 6). *HLAA*, *HLAB* and *TAP2* RNA expression showed a decreased density, suggesting suppression of antigen presentation and processing. The other immune related gene groups in tumors also showed variation, but median values were not significantly different from non-cancerous tissue. At last, we asked ourselves what impact the different immune components have on survival. Interestingly, we were unable to identify survival benefit in an adjusted Cox regression for high expression of genes involved in antigen presentation or costimulatory function (Extended Data Table 3). Genes involved in the inhibition of the immune system also were not associated with survival.

3. Discussion

We identified epigenetic and RNA expression patterns in tumor tissue from NSCLC patients, that distinguished squamous cell lung cancer from adenocarcinoma. Especially the conserved loci with hardly variation in DNA methylation between hundreds of patients were responsible for the distinction between the subtypes. Adenocarcinoma was globally more methylated than squamous cell carcinoma. Immune adaptive mechanisms have also been described such as gene hypermethylation targeting the interleukin-6/Stat3 pathway [10].

Not only methylation but also the differences in the expression values of only six genes could explain the difference between the subtypes in 97 % of patients. Involved genes were keratine 5 (*KRT5)*, tumor protein p63 (*TP63)*, *DSC3*, desmoglein 3 (*DSG3)*, calmodulin like 3 (*CALML3)*, and the miRNA *MIR205HG*. All are directly or indirectly involved in tissue morphogenesis, differentiation cell adhesion, and proliferation. The predominant isoform ΔNp63α is overexpressed in SCC and may influence tissue microenvironment by recruiting proinflammatory cells. *TP63* is commonly used in immunohistochemistry to differentiate SCC from adenocarcinoma, which supports the robustness of our analysis [11–13].

Next, we have shown that immune regulatory genes were included in regions marked by methylation probes with a ks-score > 95 %; these methylation profiles were associated with differential expression in immune modulatory genes before therapy. These genes were located in methylation regions distinguishing pretreatment NSCLC phenotypes by distribution, but were not identified by the best split analysis, indicating that genes involved in subtype morphology and immune regulation are both regulated by methylation but belong to completely distinct gene groups.

Now we have established the relation between methylation and immune expression status, we asked ourselves whether the pretreatment expression of the immune modulating gene groups of early NSCLC patients had survival consequences. We were unable to identify any

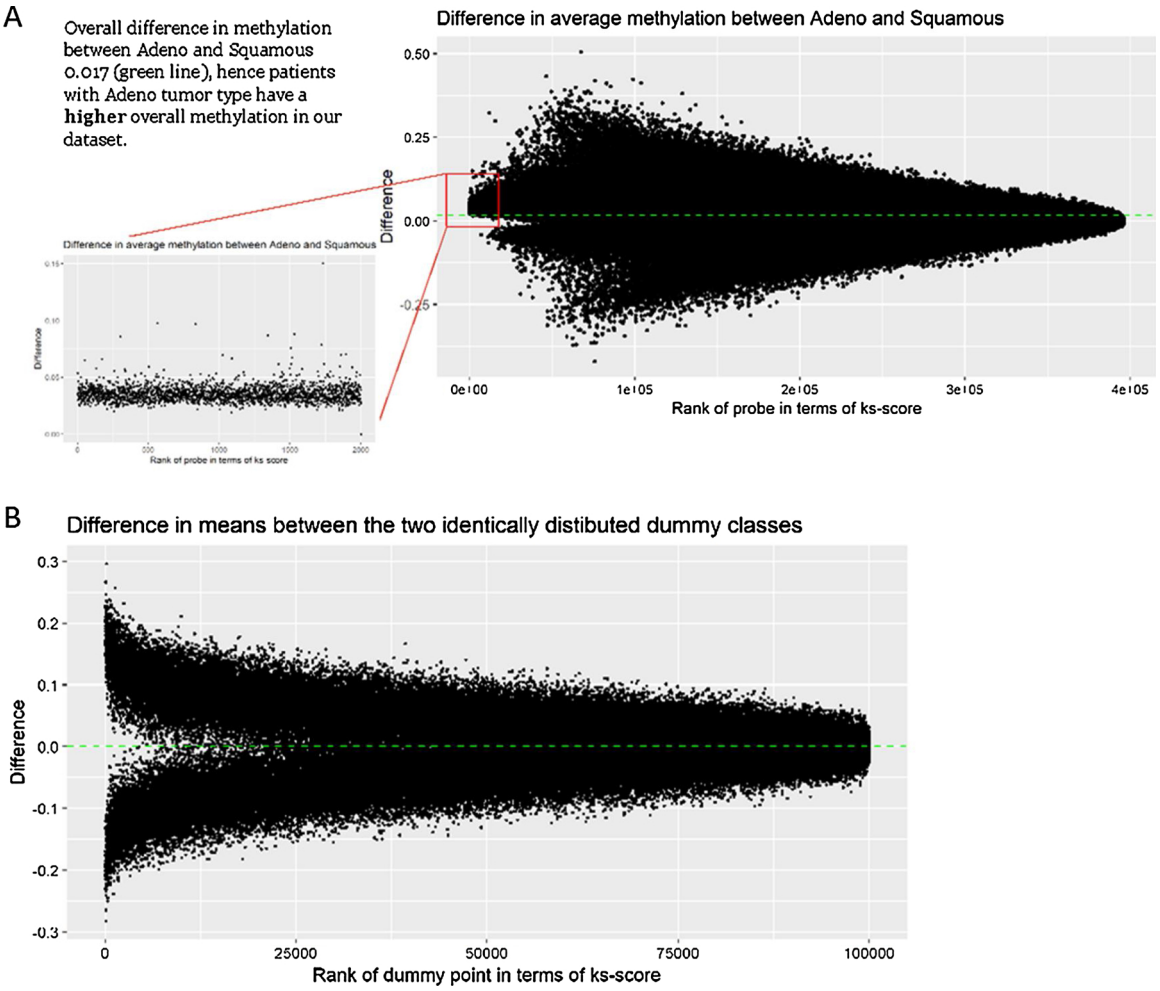

Fig. 2. Higher DNA methylation in adenocarcinoma compared to squamous cell carcinoma.
A) Adenocarcinoma contains consistently higher methylated DNA than squamous cell lung carcinoma mainly due to a relative small number of probes at conserved loci compared to B) a theoretical at random model. Overall difference in methylation is 0.017 (positive values on the y-axis indicate higher DNA methylation of adenocarcinoma, negative values indicate higher DNA methylation of squamous cell carcinoma). X-axis ranks the probes according to the ks-score for differentiation between both histological subtypes. Y-axis is the difference between the mean methylation (0 is low methylation and 1 is high methylation) between both subtypes.

survival benefit. Compared to non-cancerous tissue we observed a much broader distribution of the expression of immune modulating genes in NSCLC, while the median expression of T cell co-inhibitory, co-stimulatory, and cyto- and chemokine genes remained similar. Only the antigen presenting gene expression in NSCLC was decreased. This group consisted not only of the classical *HLAA*, *HLAB* and *HLAC* whose expression depend on methylation but also *B2M* and *TAP* genes. By multiplexed quantitative immunofluorescence loss of expression of B2M, HLA-I heavy chains and HLA-II was observed in less than 23 % of NSCLC patients [14]. Loss of *B2M* expression resulted in decreased or no cell surface expression of MHC class I, which impairs antigen presentation to cytotoxic T cells [15,16]. In melanoma loss of *B2M* and *TAP1* expression reduced overall survival when treated with ipilimumab [17]. Limiting the expression of genes involved in antigen presentation is an important mechanism of tumor cells to evade the immune system [18]. In early NSCLC tumors that have an activated immune system, extensive immune editing is present in order to fit with the tumor microenvironment, as indicated by the relative depletion of neoantigens in tumors and loss of heterozygosity in *HLA* genes [19]. This allele-specific *HLA* loss may occur in about 40 % of NSCLC patients [18]. We observed that in tumor and non-cancerous tissue antigen presenting and co-stimulatory gene expression was positively associated, suggesting that a higher expression of antigen presenting genes is associated with more inflammation. Although our analysis does not provide information on cell types, it suggests that the higher dosage of antigen presenting gene expression in (any) tissue associates with more co-stimulatory gene expressions from T cells. Overall, it may be concluded that the antigen presenting gene group harbors the main immune related defect in NSCLC patients.

Finally, our analysis revealed that higher methylation was observed in genes involved in the immune response towards endogenous retroviral sequences in adenocarcinoma compared to SCC. Disruption of methylation in both subtypes leads to different retroviral expressions. Moreover, analyzing a wide spectrum of over 2000 involved genes with the highest subtype separability revealed viral involvement, likely retroviral or transposon loci. As we know, human endogenous retroviruses are under epigenetic control and rarely expressed in normal tissue [20,21]. Hypomethylation of the LINE family member L1 occurs in multiple solid cancers and cell lines [22,23]. Lung squamous cell carcinoma has elevated ERVH-5 and other RNA derived endogenous retrovirus expression that were associated with low cytolytic activity [24]. We observed that IRF3, NF-κB and STAT pathways are critical in the production of type I interferons downstream of pathogen recognition receptors. They detect viral RNA and DNA [25]. The genes *SRPK1* and *TRIM4* are found to regulate these virus-induced IFN induction pathways [26,27]. This provides further molecular evidence of the presumed importance of (retro)viral infection in predominantly squamous cell carcinoma as previously observed in squamous cell carcinomas that

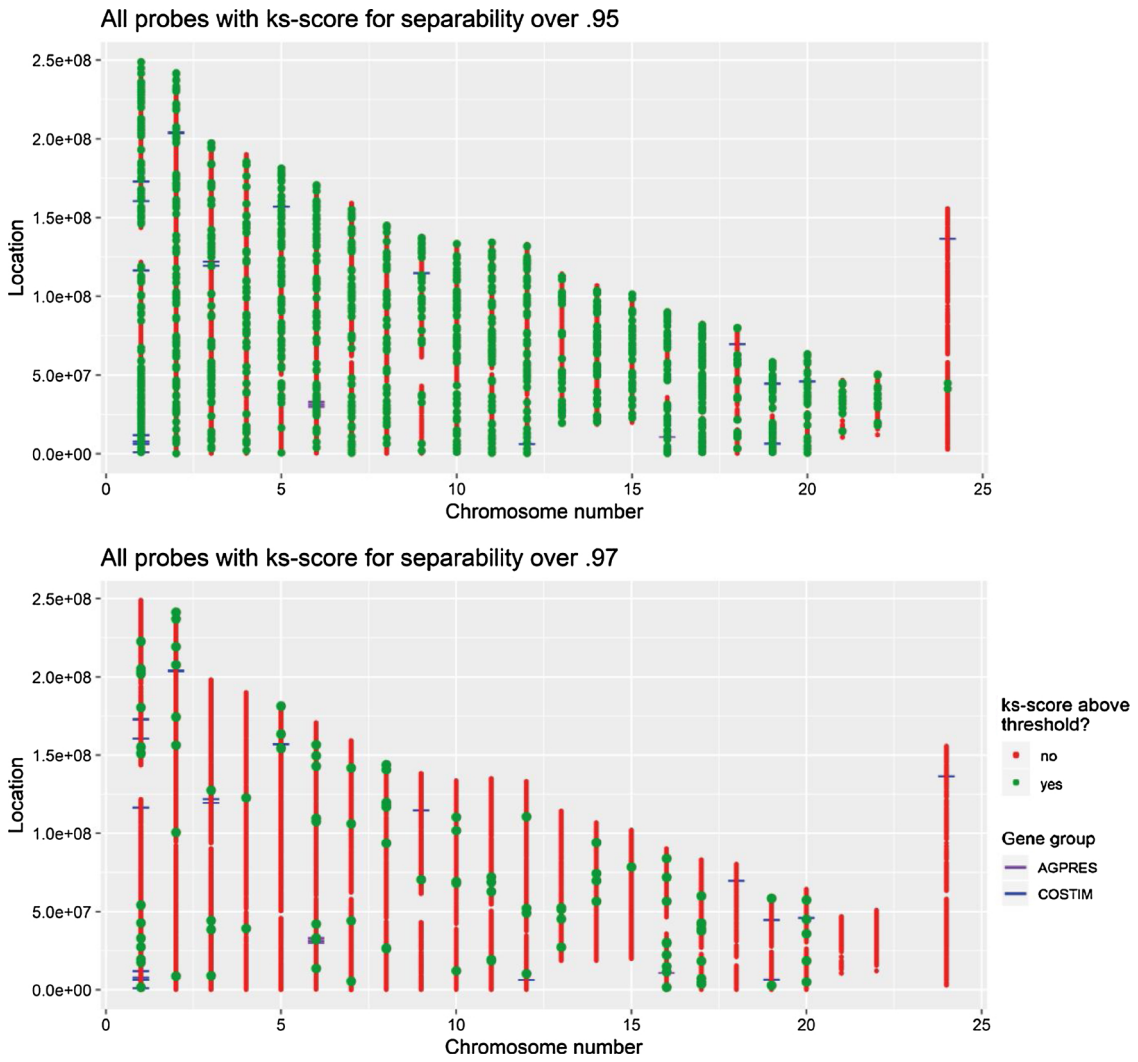

Fig. 3. Location of methylation probes along 23 chromosomes that separates adenocarcinoma from squamous cell lung carcinoma at ks ≥ 0.95 and ks ≥ 0.97 level. Methylation pattern for each chromosome characterized by their individual probes A) with ks-score for separability between histologic subtypes over 95 % are evenly distributed over chromosomes with an exception for the x-chromosome. B) Probes with ks-score over 97 % show the conserved methylated areas that preserve the difference between subtypes. Mostly they are related to CpG islands located along chromosomes. X-axis and y-axis refer to respectively the chromosome number and individual probe localization on the chromosome according to ks-score for separability. Green dots represent differential probes. Antigen presentation and costimulation genes are flagged for chromosome location. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

contained viral DNA [28]. Importantly, a significant proportion of the differentiating methylation probes suppresses viral and retroviral associated genes. Further investigation is needed into the exact nature of those repetitive areas that are hyper- or hypomethylated.

All studies have limitations. Importantly, the tumor samples have varying tumor content (at least 40 %) tough, as shown by stratification, this had no consequences for our findings. In order to perform our enrichment analysis of the epigenetic background across NSCLC phenotypes on a gene level methylation signals were averaged. Although this approach resulted in relevant and consistent findings, it may lead to inevitable loss of information as its effect varies across different gene regions. For instance, hypermethylation of high density CpG regions has been recognized to strongly associate with gene expression regulation [29]. Lastly, splicing variants and small cumulative effects within several genes in the same pathway have not taken into consideration in the RNA expression analysis. Alternative splicing may have a functional impact and is increased in cancer compared to normal tissue [29].

Together these results show that NSCLC phenotypes are largely determined by epigenetic regulation of a small conserved group of genes, involved in extracellular matrix and cell structure. Methylation controls immune related genes – also those involved in endogenous retroviral sequences - that show a larger expression diversity in tumor than in non-cancerous tissue. Decreased expression of genes involved in antigen presentation are the main immune related defect in NSCLC, highlighting their importance for immune invasion by the tumor.

4. Methods

4.1. Study cohort and data acquisition

Patients with treatment naive NSCLC, adenocarcinoma and SCC, whose DNA methylation and RNA expression data from the resected tumor was available in the public domain, were selected from two different profiling platforms (RNA sequencing resulting in 60,483 mRNA expression values and methylation profiling by Infinium HM450 platform resulting in 485,577 DNA methylation β-values) at the TCGA Research Network (http://cancergenome.nih.gov/). Details on methods and data generation of the RNA sequencing and DNA methylation can be found in the original TCGA landmark papers [3,4] Duplicate

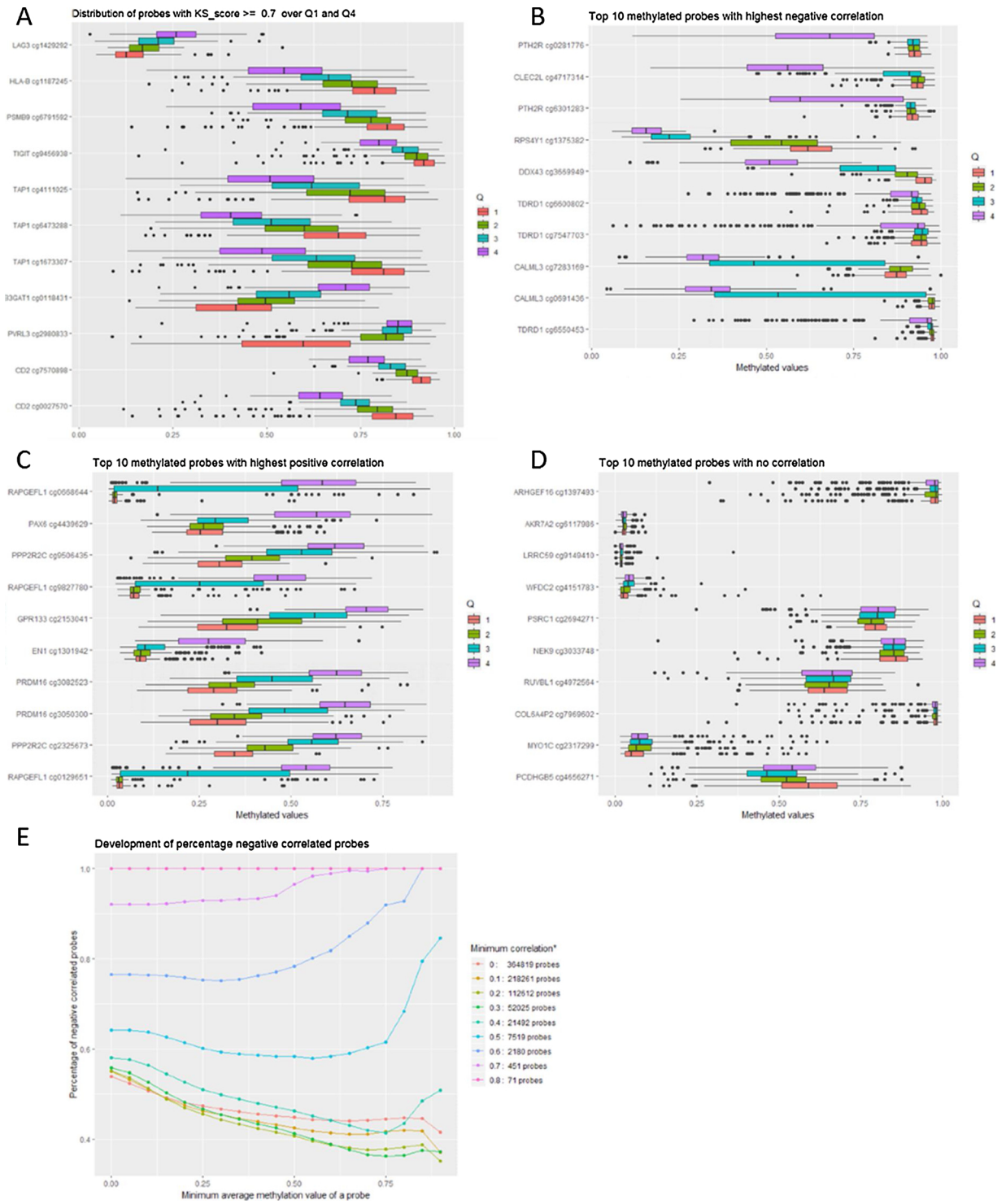

Fig. 4. Relation between DNA methylation for immune modulating genes determined by one probe and gene expression at four quartile levels.
A) Immune modulating genes identified by probes with ks-score ≥ 70 % (adenocarcinoma vs. SCC) show an inverse relationship between expression and methylation for most genes. In this example, LAG3 and B3GAT1 show the opposite expression effect at low and moderate methylation, respectively. Examples are shown of probes with B) the highest negative correlation, C) highest positive correlation and D) no correlation between DNA methylation and RNA expression. E) The percentage of methylation probes that are inversely correlated with RNA expression depends on the cut-off of the correlation between methylation and gene expression of a probe and minimal average methylation.

samples, those with missing histological diagnosis, and those with disease recurrences were removed. In total, 1024 unique NSCLC patients with tumor tissue were selected of whom 154 patients provided additional normal tissue. Patients with normal tissue provided 108 samples for the normal RNA expression dataset and 74 samples for the normal DNA methylation dataset; 28 patients provided samples for both methods. The tumor RNA expression dataset consisted of 1014 tumor samples (513 adenocarcinoma and 501 SCC) and the tumor methylation dataset consisted of 828 tumor samples (458 adenocarcinoma and 370 SCC). We extracted clinical and pathological data on age, gender, histology, stage of disease, tumor cell percentage and survival calculated from time of diagnosis to time of death or last follow-up (Extended

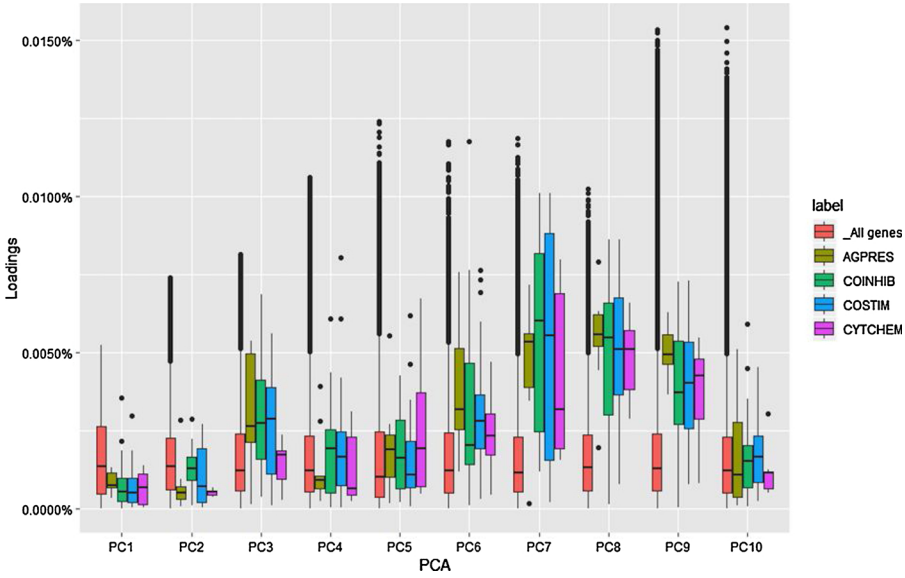

Fig. 5. Four immune modulatory gene groups as compared with all gene expressions in non-small cell lung cancers shows two clusters of increase (PC3 and PC6-9).
Four main immune modulatory gene groups were distinguished, involved in T cell antigen presentation (AGPRES), T cell co-inhibitory (COINHIB), T cell co-stimulator (COSTIM) and T cell cytokines/chemokines (CYTOCHEM). The influence of these gene groups were investigated on the individual principle components. The y-axis represents the loading of a gene in the DNA expression dataset to a principle component. The red boxes indicate all genes in the DNA expression dataset, whereas the other boxes represent the genes of selected immune modulating gene groups. All immune modulating gene groups are most pronounced in PC7, PC8 and PC9. The midline in the boxplot is the median of data in that component, with the lower and upper limits of the box being the first and third quartile, respectively. By default, the whiskers will extend up to 1.5 times the interquartile range from the top (bottom) of the box. If there are any data beyond that distance, they are represented individually as black dots ('outliers'). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Data Table 4). The dataset was analyzed during a hackathon session, in which data scientists in collaboration with physicians competed to create a “functioning” product by the end of the 3-day event.

4.2. Data curation and statistical methods

All datasets were filtered and curated for non-significantly associated features after preprocessing of the files. ComBat, an empirical Bayes location/scaling method was applied to rule out potential cohort bias in the RNA expression data as a consequence of different study sites and laboratories, whereas BEclear was used in the DNA methylation data (Supplementary Information). As co-variant we used gender, because this factor has a high variance over the batches and its value is known for all samples. We started with principal component analysis (PCA) to discern underlying structure of the database, e.g. the total gene expression versus the immune modulating gene groups expressions.

For RNA expression and DNA methylation data we used non-parametric tests. The separation of the cancer types was compared before and after bias correction with both the Kolmogorov-Smirnov test and Mann-Whitney *U* test The Kolmogorov-Smirnov tests provides a ks-score, which determines whether the given distributions of two groups are the same or different with a probability of 1-ks-score (Supplementary Information). After we determined differences in distributions of methylation probes, an algorithm was developed using separate cut-off values for DNA methylation and RNA expression to identify the most predictive genes to classify the NSCLC subtypes. This best split analysis determined whether a cut-off value could provide a split of at least 85 % of patients into the correct subtype with a certainty of more than 80 %, starting with probes or genes that had the highest differences (fold change). Differences between tumor subtypes based on DNA methylation β value had to be at least 0.1 to increase the probability of biological relevance. Loci with the largest differences for both DNA methylation and RNA expression respectively were determined after the annotating the probes into corresponding genes, an overlap in genes of both lists was established for biological interpretation.

Different immune modulatory genes were selected and grouped according to their function (Extended Data Table 5). These include co-stimulatory genes (COSTIM), co-inhibitory genes (COINHIB), antigen presenting genes (AGPRES) and immune modulatory/ inflammatory cytokines and chemokines (CYTCHEM). The co-stimulatory immune modulatory gene group included genes that are known to be expressed in tumor cells (e.g. *ICOSL, OX40 L, SLAM*), as reviewed by Chen and Flies [30]. Similarly, co-inhibitory genes were included in the analysis (e.g. *VTCN1*, *CD113, CD48*). *HLAE* was included for its protein function as inhibitor ligand for immunocompetent (NK) cells [31,32]. For antigen presentation, genes were selected involved in antigen presentation (e.g. classical HLA) and genes involved in antigen processing (e.g. *TAP1, CIITA, HLAA*) were selected for inclusion in the antigen presentation genes group [33]. Genes coding for cytokines and chemokines (e.g. *IL10, IDO, IFNG*) were selected based on their implication in immune tolerance of cancer through pleiotropic effects in immune regulation and inflammation [34–36].

Gene densities of all AGPRES genes were calculated within R using a kernel density estimate from the distribution of RNA expression of NSCLC and non-cancerous tissue.

To study the relationship between expression of immune related gene groups (AGPRES, COSTIM, COINHIB, CYTCHEM) and overall survival, a multivariate Cox regression analysis was used with age, gender, smoking (pack years), tumor type and stage of disease as covariates (patient factors with $p < 0.1$ from univariate analysis included). The expression of pretreatment immune related gene groups was used as a categorical variable with two levels divided by the median (high and low overall expression of all involved genes). Hazard ratios (HR) and 95 % confidence intervals (CI) are reported.

To investigate the biological pathways, Ingenuity Pathway Analysis (Qiagen, Hilden, Germany) was used to perform gene enrichment analyses on these gene lists.

Transparency document

The Transparency document associated with this article can be found in the online version.

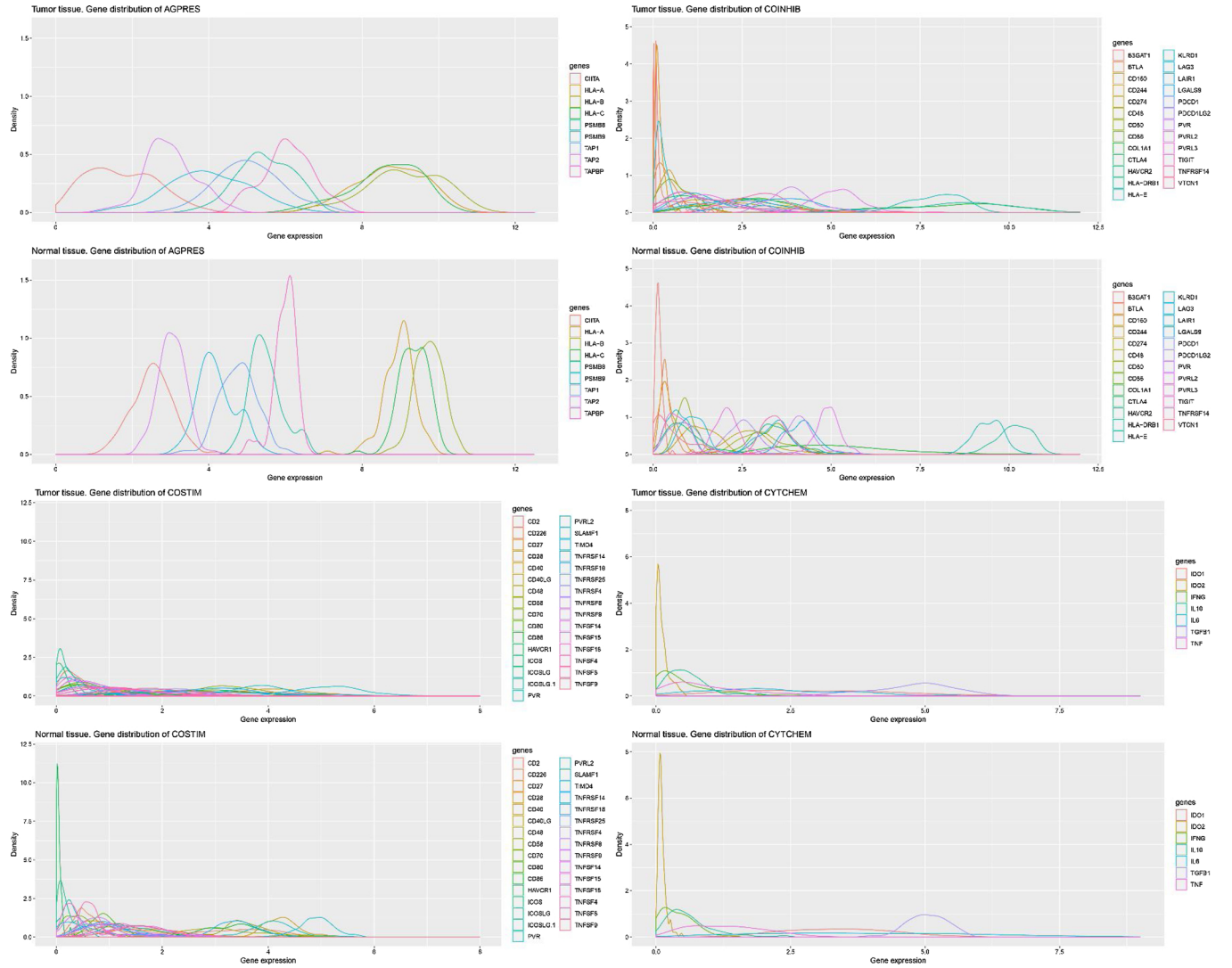

Fig. 6. RNA expression of immune modulating genes in tumor and their matched non-cancerous tissue.
Density distribution of immune modulating gene expressions in 106 NSCLC tumors and their matched non-cancerous tissue. The antigen presenting gene expressions show a different density distribution compared to non-cancerous expressions, while the density of coinhibitory, costimulatory, and cyto- and chemokine gene expressions were largely similar.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.lungcan.2020.06.008.

References

[1] H.J.M. Groen, T.J.N. Hiltermann, Air pollution and adenocarcinoma in never-smokers, J. Thorac. Oncol. 14 (2019) 761–763, https://doi.org/10.1016/j.jtho.2019.02.007.
[2] T. Sato, E. Arai, T. Kohno, et al., Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: its correlation with smoking history and chronic obstructive pulmonary disease, Int. J. Cancer 135 (2014) 319–334, https://doi.org/10.1002/ijc.28684.
[3] P.S. Hammerman, M.S. Lawrence, D. Voet, et al., Comprehensive genomic characterization of squamous cell lung cancers, Nature 489 (2012) 519–525, https://doi.org/10.1038/nature11404.
[4] E.A. Collisson, J.D. Campbell, A.N. Brooks, et al., Comprehensive molecular profiling of lung adenocarcinoma, Nature 511 (2014) 543–550, https://doi.org/10.1038/nature13385.
[5] K.A. Hoadley, C. Yau, D.M. Wolf, et al., Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin, Cell 158 (2014) 929–944, https://doi.org/10.1016/j.cell.2014.06.049.
[6] F. Chen, Y. Zhang, E. Parra, et al., Multiplatform-based molecular subtypes of non-small-cell lung cancer, Oncogene 36 (2017) 1384–1393, https://doi.org/10.1038/onc.2016.303.
[7] C. Moison, C. Senamaud-Beaufort, L. Fourrière, et al., DNA methylation associated with polycomb repression in retinoic acid receptor β silencing, FASEB J. 27 (2013) 1468–1478, https://doi.org/10.1096/fj.12-210971.
[8] A.J. Gentles, S.V. Bratman, L.J. Lee, et al., Integrating tumor and stromal gene expression signatures with clinical indices for survival stratification of early-stage non-small cell lung cancer, J. Natl. Cancer Inst. 107 (2015) 1–11, https://doi.org/10.1093/jnci/djv211.
[9] L. Horn, D.R. Spigel, E.E. Vokes, et al., Nivolumab versus docetaxel in previously treated patients with advanced non–small-cell lung cancer: two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057), J. Clin. Oncol. 35 (2017) 3924–3933, https://doi.org/10.1200/jco.2017.74.3062.
[10] X. Wang, Y. Wang, G. Xiao, et al., Hypermethylated in cancer 1(HIC1) suppresses non-small cell lung cancer progression by targeting interleukin-6/Stat3 pathway, Oncotarget 7 (2016) 30350–30364, https://doi.org/10.18632/oncotarget.8734.
[11] C.A. Wright, F.R.C. Path, M. Van Der Burg, et al., Diagnosing mycobacterial lymphadenitis in children using fine needle aspiration biopsy : cytomorphology, ZN staining and autofluorescence — making more of less, Diagn. Cytopathol. 36 (2008) 245–251, https://doi.org/10.1002/dc.
[12] W.D. Travis, E. Brambilla, M. Noguchi, et al., International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma, Physiol. Behav. 176 (2019) 139–148, https://doi.org/10.1016/j.physbeh.2017.03.040.

[13] A.G. Nicholson, D. Gonzalez, P. Shah, et al., Refining the diagnosis and EGFR status of non-small cell lung carcinoma in biopsy and cytologic material, using a panel of Mucin staining, TTF-1, cytokeratin 5/6, and P63, and EGFR mutation analysis, J. Thorac. Oncol. 5 (2010) 436–441, https://doi.org/10.1097/JTO.0b013e3181c6ed9b.
[14] I. Datar, K.A. Villarroel-Espindola, Franz Henick, Brian S. Syrigos, Konstantinos N. Toki, Maria Rimm, David L. Ferrone, Soldano Herbst, Roy S. Schalper, J. Clin. Oncol. 36 (2018), https://doi.org/10.1200/JCO.2018.36.15_suppl.12015.
[15] P. Leone, E.C. Shin, F. Perosa, et al., MHC class i antigen processing and presenting machinery: organization, function, and defects in tumor cells, J. Natl. Cancer Inst. 105 (2013) 1172–1187, https://doi.org/10.1093/jnci/djt184.
[16] J.M. Zaretsky, A. Garcia-Diaz, D.S. Shin, et al., Mutations associated with acquired resistance to PD-1 blockade in melanoma, N. Engl. J. Med. 375 (2016) 819–829, https://doi.org/10.1056/NEJMoa1604958.
[17] S.J. Patel, N.E. Sanjana, R.J. Kishton, et al., Identification of essential genes for cancer immunotherapy Shashank, Nature 548 (2015) 537–542, https://doi.org/10.1038/ncomms5930.Genetic.
[18] N. McGranahan, R. Rosenthal, C.T. Hiley, et al., Allele-specific HLA loss and immune escape in lung cancer evolution, Cell 171 (2017) 1259–1271, https://doi.org/10.1016/j.cell.2017.10.001 e11.
[19] R. Rosenthal, E.L. Cadieux, R. Salgado, et al., Neoantigen-directed immune escape in lung cancer evolution, Nature 567 (2019) 479–485, https://doi.org/10.1038/s41586-019-1032-7.
[20] A.S. Attermann, A.M. Bjerregaard, S.K. Saini, et al., Human endogenous retroviruses and their implication for immunotherapeutics of cancer, Ann. Oncol. 29 (2018) 2183–2191, https://doi.org/10.1093/annonc/mdy413.
[21] S.R. Richardson, A.J. Doucet, H.C. Kopera, et al., The influence of LINE-1 and SINE retrotransposons on mammalian genomes, Mob. DNA III 3 (2015) 1165–1208, https://doi.org/10.1128/microbiolspec.mdna3-0061-2014.
[22] E.M. Wolff, H.M. Byun, H.F. Han, et al., Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer, PLoS Genet. 6 (2010), https://doi.org/10.1371/journal.pgen.1000917.
[23] C. Phokaew, S. Kowudtitham, K. Subbalekha, et al., LINE-1 methylation patterns of different loci in normal and cancerous cells, Nucleic Acids Res. 36 (2008) 5704–5712, https://doi.org/10.1093/nar/gkn571.
[24] M.S. Rooney, S.A. Shukla, C.J. Wu, et al., Molecular and genetic properties of tumors associated with local immune cytolytic activity, Cell 160 (2015), https://doi.org/10.1016/j.physbeh.2017.03.040.
[25] C.A. Jefferies, Regulating IRFs in IFN driven disease, Front. Immunol. 10 (2019) 1–15, https://doi.org/10.3389/fimmu.2019.00325.
[26] L. Nousiainen, M. Sillanpää, M. Jiang, et al., Human kinome analysis reveals novel kinases contributing to virus infection and retinoic-acid inducible gene I-induced type i and type III IFN gene expression, Innate Immun. 19 (2013) 516–530, https://doi.org/10.1177/1753425912473345.
[27] J. Yan, Q. Li, A.P. Mao, et al., TRIM4 modulates type i interferon induction and cellular antiviral response by targeting RIG-I for K63-linked ubiquitination, J. Mol. Cell Biol. 6 (2014) 154–163, https://doi.org/10.1093/jmcb/mju005.
[28] L.A. Robinson, C.J. Jaing, C. Pierce Campbell, et al., Molecular evidence of viral DNA in non-small cell lung cancer and non-neoplastic lung, Br. J. Cancer 115 (2016) 497–504, https://doi.org/10.1038/bjc.2016.213.
[29] M. Moarii, V. Boeva, J.P. Vert, F. Reyal, Changes in correlation between promoter methylation and gene expression in cancer, BMC Genomics 16 (2015) 1–14, https://doi.org/10.1186/s12864-015-1994-2.
[30] L. Chen, D.B. Flies, Molecular mechanisms of T cell co-stimulation and co-inhibition, Nat. Rev. Immunol. 13 (2013) 227–242, https://doi.org/10.1038/nri3405.
[31] N. Lee, M. Llano, M. Carretero, et al., HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 5199–5204, https://doi.org/10.1073/pnas.95.9.5199.
[32] J. Eugène, N. Jouand, K. Ducoin, et al., The inhibitory receptor CD94/NKG2A on CD8+ tumor-infiltrating lymphocytes in colorectal cancer: a promising new druggable immune checkpoint in the context of HLAE/β2m overexpression, Mod. Pathol. 33 (3) (2019) 468–482, https://doi.org/10.1038/s41379-019-0322-9.
[33] K.S. Kobayashi, P.J. Van Den Elsen, NLRC5: a key regulator of MHC class I-dependent immune responses, Nat. Rev. Immunol. 12 (2012) 813–820, https://doi.org/10.1038/nri3339.
[34] J.M. Vahl, J. Friedrich, S. Mittler, et al., Interleukin-10-regulated tumour tolerance in non-small cell lung cancer, Br. J. Cancer 117 (2017) 1644–1655, https://doi.org/10.1038/bjc.2017.336.
[35] D.H. Munn, M.D. Sharma, A.L. Mellor, Ligation of B7-1/B7-2 by human CD4 + T cells triggers indoleamine 2,3-dioxygenase activity in dendritic cells, J. Immunol. 172 (2004) 4100–4110, https://doi.org/10.4049/jimmunol.172.7.4100.
[36] W.Z. Nisha Nagarsheth, M.S. Wicha, Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy, Nat. Rev. Immunol. 17 (2017) 559–572, https://doi.org/10.1016/j.physbeh.2017.03.040.